The EMBO Journal vol.9 no.12 pp.3807 – 3814, 1990

Three-dimensional structure determination of an anti-2-phenyloxazolone antibody: the role of somatic mutation and heavy/light chain pairing in the maturation of an immune response

Pedro M.Alzari, Silvia Spinelli, Roy A.Mariuzza, Ginette Boulot, Roberto J.Poljak, John M.Jarvis[1] and César Milstein[1]

Unité d'Immunologie Structurale (URA 359 CNRS), Département d'Immunologie, Institut Pasteur, 75724 Paris Cedex 15, France and the [1]Medical Research Council Laboratory of Molecular Biology, Hills Road, Cambridge, CB2 2QH, UK

Communicated by C.Milstein

The three-dimensional structure of the Fab fragment of an anti-2-phenyloxazolone monoclonal antibody (NQ10/12.5) in its native and complexed forms has been determined at 2.8 and 3.0 Å resolution, respectively. Identification of hapten-contacting residues has allowed us to evaluate the contribution of individual somatic point mutations to maturation of the immune response. In particular, amino acid residues 34 and 36 of the light chain, which are frequently mutated in antibodies with increased affinity for 2-phenyloxazolone, are shown to interact directly with the hapten. We propose that the strict maintenance of certain amino acid sequences at the potentially highly variable V_L-J_L and V_H-D-J_H junctions observed among anti-2-phenyloxazolone antibodies is due largely to structural constraints related to antigen recognition. Finally, the three-dimensional model of NQ10/12.5, which uses the typical light chain of primary response anti-2-phenyloxazolone antibodies but a different heavy chain, allows an understanding of how, by preserving key contact residues, a given heavy chain may be replaced by another, apparently unrelated one, without loss of hapten binding activity and why the V_xOx1 germline gene is so frequently selected amongst the other known members of this family.

Key words: anti-phenyloxazolone antibody/crystal structure/hapten–Fab interactions/maturation of an immune response/somatic mutations

Introduction

The principal genetic mechanisms responsible for the generation of sequence diversity in the genes encoding the heavy (H) and light (L) chains of immunoglobulin molecules now seem to have been identified (Tonegawa, 1983). These include (i) combinatorial diversification whereby three sets of H chain gene segments, V_H, D and J_H, and two sets of L chain gene segments, V_L and J_L, must rearrange to generate functional variable regions; (ii) imprecise joining of these gene segments at the V_H-D, $D-J_H$ and V_L-J_L junctions; and (iii) somatic point mutation by which base changes are introduced throughout the sequences encoding H and L chains.

Extensive sequence analysis of antibodies elicited by a number of model haptens, including 2-phenyloxazolone (phOx; Berek *et al.*, 1985), phenylarsonate (Wysocki *et al.*, 1986), nitrophenylacetyl (Bothwell *et al.*, 1981) and phosphorylcholine (Crews *et al.*, 1981), has been performed in order to clarify the way in which the above mechanisms operate during the onset and maturation of an immune response. In the case of 2-phenyloxazolone, the initial response is dominated by very similar or identical antibodies expressing a single L/H chain combination encoded by a pair of germline genes, V_xOx1 and V_HOx1, respectively (Kaartinen *et al.*, 1983). Later in the response, the same germline gene combination is still used, but the sequences begin to diverge as affinity for hapten increases, indicating the importance of somatic point mutations in maturation of the response (Griffiths *et al.*, 1984). In the secondary response to phOx, antibodies encoded by new germline gene combinations become dominant; these display generally higher affinities than the original V_xOx1/V_HOx1 antibodies (Berek *et al.*, 1985, 1987). Thus, while somatic mutation is the major source of improved antibodies, affinity maturation is also achieved through the recruitment of alternative germline gene combinations (repertoire shift).

The observations made on the maturation of the immune response to phOx are generally consistent with those made in other hapten systems. Thus, primary antibodies to nitrophenylacetyl have λ chains and are largely unmutated, whereas hyperimmune antibodies have κ chains, are somatically mutated, and display increased affinity for the hapten (Jack *et al.*, 1977; Reth *et al.*, 1978; Bothwell *et al.*, 1981; Sablitzky *et al.*, 1985). In the anti-phenylarsonate response of strain A mice, a V_H region that is present in only a minority of primary response antibodies, but which bestows the highest affinity for hapten, dominates the secondary response and antibodies bearing mutated derivatives of this V_H region bind phenylarsonate with even higher affinity (Wysocki *et al.*, 1986; Manser *et al.*, 1987).

The maturation of an immune response raises a number of important structural questions. What proportion of the somatic mutations accumulated during maturation are functionally significant (i.e. affect residues involved in antigen binding, either directly or indirectly), as opposed to simply reflecting a background mutation rate? An answer to this question is required to assess properly the extent to which antigen favours the selective expansion of clones producing antibodies bearing mutations which impart higher affinity. What are the structural constraints responsible for the restrictions in amino acid sequences at the potentially highly variable V_L-J_L and V_H-D-J_H junctions often observed among antibodies of the same specificity? How is it possible to pair a given L chain with different H chains, or vice versa, and still preserve a functional antigen-combining site? In such situations, is the basic geometry of the antibody active site likely to be maintained, at least for regions in contact with antigen or, on the other hand, is it

necessary to postulate fundamentally altered modes of antigen binding? Finally, is the recruitment of alternative germline gene combinations observed in the secondary response a consequence of the difficulty in further improving the affinity of the original V_xOx1/V_HOx1 combination? (Berek *et al.*, 1985; Manser *et al.*, 1987).

To begin to address these problems, we have determined the three-dimensional structure of a member of the anti-phOx family of antibodies, NQ10/12.5, in its native and hapten-bound forms. This secondary response antibody is composed of an L chain most likely originating through several point mutations from the germ-line gene V_xOx1 and of a H chain belonging to the V_H-MOPC21 gene family (Berek *et al.*, 1985). An analysis of the structure of NQ10/12.5, in the context of the extensive sequence information available in this system, has allowed us to identify those amino acid residues implicated in hapten binding and to thereby evaluate the contribution of individual point mutations to affinity maturation of the anti-phOx response. In addition, the three-dimensional model of NQ10/12.5 allows an understanding of how, by retaining key contact residues, a specific mode of hapten binding may be preserved, even following substitution of a given H (or L) chain by another, apparently unrelated one.

Table I. Refinement data for FabNQ10/12.5

R-factor[a]	0.182
Data resolution	7–2.8 Å
Number of observations with I > 2 σ (I)	10290
rms deviations from ideal bond lengths	0.018 Å
rms deviations from ideal bond angles	3.8°
rms deviations form ideal dihedral angles	28.9°

[a] $R = \sum_h ||F_o| - |F_c|| / \sum_h |F_o|$

Results

The three-dimensional structure of FabNQ10/12.5 was initially solved at 6 Å resolution using SIR phases from a uranyl acetate heavy atom derivative. Subsequently its atomic coordinates were refined at 2.8 Å to an *R*-value of 0.18 by alternative cycles of least-squares refinement and manual model building (see Materials and methods). The final results of the crystallographic refinement are summarized in Table I.

The quarternary structure of FabNQ10/12.5 is similar to that of other Fab fragments whose three-dimensional structure has been determined by X-ray diffraction (reviewed by Amzel and Poljak, 1979; Alzari *et al.*, 1988; Davies *et al.*, 1988). The root-mean-square (rms) positional differences after least-squares superposition of selected Cα coordinates between the V regions of Fab NQ10/12.5 and other Fabs are in the range of 0.6–1.0 Å.

The molecule is in a rather extended conformation, with an 'elbow' angle made by the pseudo two-fold axes relating V_H to V_L and C_H1 to C_L of ~159°. The rotation–translation operations that optimize the superposition of conserved Cα positions of the V_H and V_L subunits are 174.2° and 0.40 Å, respectively. These values are similar to those observed in other Fabs, indicating a global conservation of V_H–V_L association among known Fab structures (Lascombe *et al.*, 1989). The numerical values for the corresponding operations relating C_H1 to C_L are 173.1° and 1.65 Å.

The molecules are linearly arranged in a head to tail configuration; adjacent lines intersect near the crystallographic two-fold axis and form an angle close to 90° (Figure 1). This gives rise to extended intermolecular contacts in this region, which preclude solvent exposure of a large fraction of the complementarity determining regions' (CDR) external surface. As a consequence of crystal packing, an intermolecular β-sheet is observed between two V_H subunits related by the two-fold axis. The outermost

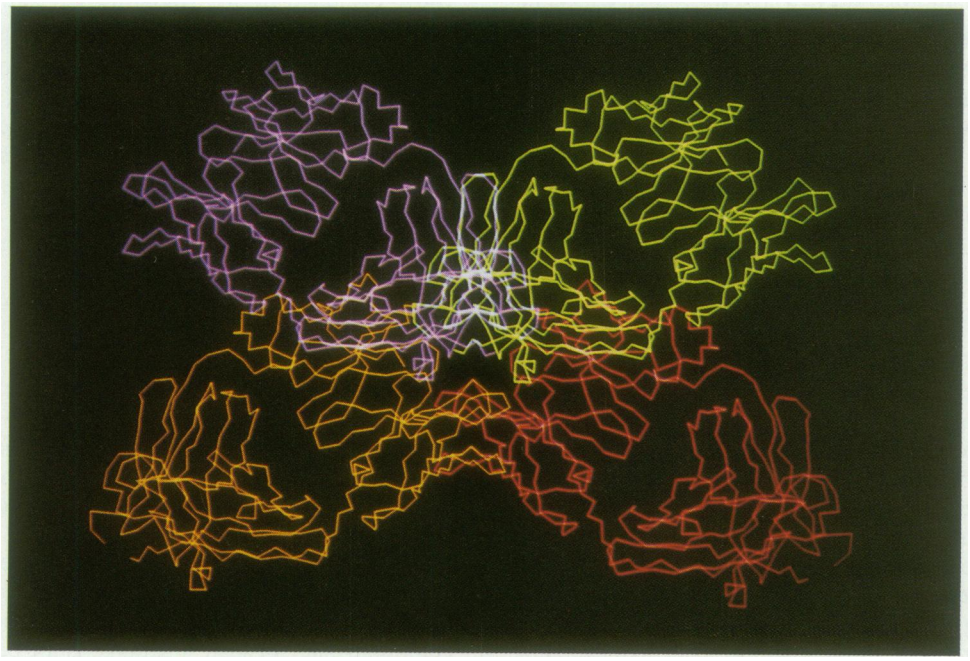

Fig. 1. Projection of the α-carbon skeleton of four Fab molecules showing the head-to-tail arrangement; the crystallographic two-fold axis is contained in the plane of the figure. The intermolecular β-sheet, formed by two V_H subunits related by the symmetry axis, can be seen in this view between the red and orange Fab backbone models.

strands of the internal sheets from these two subunits are hydrogen-bonded to each other, resulting in an extended β-sheet. This type of intermolecular interaction has already been observed for two L chain dimers, Mcg and Loc (Ely *et al.*, 1978; Chang *et al.*, 1985), though in those structures the interactions involved two constant subunits.

Well-defined electron density was observed for all the CDRs and for residues involved in V_H-V_L interactions (Figure 2). Only the side chain of V_H Tyr96 (numbering according to Kabat *et al.*, 1987), located in the centre of CDR3, is not visible in the electron density map. The conformation of this loop implies that the unobserved side chain is directed towards the solvent region. An electron density peak is observed in the unliganded Fab structure, in the cavity occupied by the hapten in the complex (Figure 2). This peak can be assigned to a solvent molecule which is hydrogen-bonded to V_L Tyr32. Analysis of the $2F_{obs}-F_{calc}$ map for the phOx−Fab complex showed the largest, elongated peak, which we assigned to the hapten, to be well buried in the central region of the antigen binding site. The accessible surface area excluded from the solvent due to the hapten−Fab interactions is ~400 Å^2 (~230 Å^2 from the Fab and ~170 Å^2 from the hapten). Most of the complex-occluded Fab surface area (Table II) corresponds to an internal cavity already existent in the binding site of the native Fab crystal structure (Figure 2a).

Seven L chain residues (two from CDR1, three from CDR3, one from FR2 and one from FR4), and seven H chain residues (one from CDR1, three from CDR3, one from FR2, one from FR3 and one from FR4) make at least one contact with the hapten (distance <4 Å; see Table II and Figure 3).

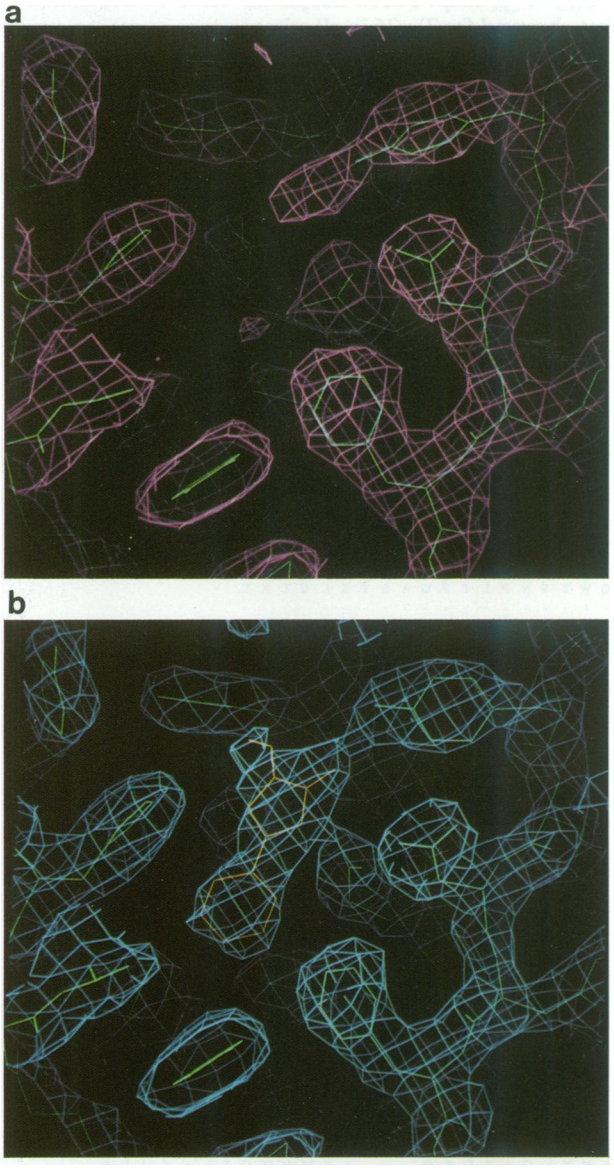

Fig. 2. Electron density map, calculated with coefficients ($2F_o-F_c$), corresponding to the hapten binding site in the free (**a**) and liganded (**b**) crystal structures. The three antigen-contacting residues from the CDR1-FR2 segments from the L chain (Tyr L32, Asn L34 and Phe L36) are seen to the right of the figure, while His H35 can be seen on the left. The solvent exposed CDR surface is at the top of the figure, and the CDR3 of the L chain is just behind the binding cavity on this projection (the side chain of Gln L89 is visible in the background).

Table II. Fab residues in contact with the hapten

Position	Amino acid	Segment	Shortest distance (Å)	Complex-occluded surface area[a] (Å^2)
L32	Tyr	CDR1	2.7	13
L34	Asn	CDR1	3.6	10
L36	Phe	FR2	3.4	12
L89	Gln	CDR3	2.9	23
L91	Trp	CDR3	3.9	22
L96	Leu	CDR3	3.9	21
L98	Phe	FR4	3.4	10
H35	His	CDR1	3.7	25
H37	Val	FR2	3.8	10
H93	Ala	FR3	3.3	13
H95	Asp	CDR3	3.1	11
H96	Tyr	CDR3	3.1	35
H97	Gly	CDR3	3.7	17
H103	Trp	FR4	3.7	8

[a]Surface area calculated with MS program (Connolly, 1983) using standard van der Waals radii and a probe sphere of radius 1.4 Å. Residues of Fab NQ10/12.5 having at least one interatomic contact (distance ≤4 Å) with the phenyloxazolone moiety of the hapten. This interaction pattern may change slightly with higher resolution refinement of the model.

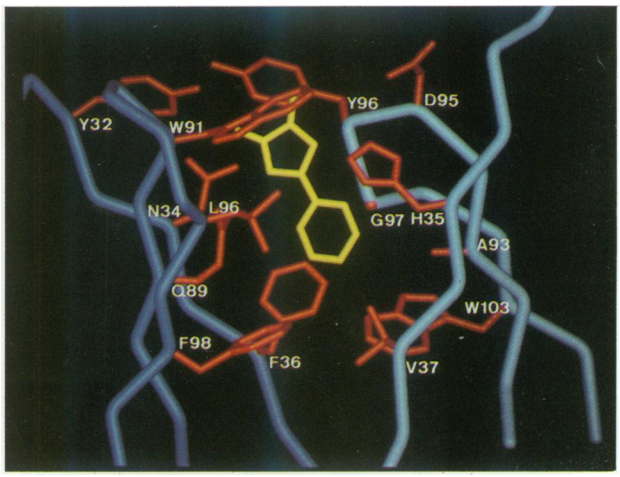

Fig. 3. Schematic view of Fab residues (side chains in red) which are in contact with the hapten (yellow). The Cα trace of the L chain, in dark blue, is shown to the left of the figure, the H chain (light blue) to the right. The second CDR from both chains has been removed to give a clear view of the binding site.

Electron density connections with the hapten are seen for two of them, V_L Tyr32 and Gln89, indicating especially close interactions. Glycine 97 of V_H CDR3 is structurally important because any side chain atom at this position would interfere with the hapten; moreover the phi,psi angles adopted by this residue (unacceptable for any other type of residue) are probably needed for closing the unusually short CDR loop. Interestingly, the same occurs in HyHEL-10, whose V_H CDR3 is of the same length and which also has a Gly at this position with main chain torsion angles similar to those observed in NQ10/12.5 (Padlan *et al.*, 1989).

V_L Tyr32 seems to be hydrogen-bonded to the O atom of the oxazolone ring. It is conserved in all described anti-phOx antibodies from the primary and secondary responses with the sole exception of NQ10/12.4.7 (V_x-ars) which has a serine at this position and does not belong to the group expressing highest affinity for the hapten (Berek *et al.*, 1985). In the whole database of kappa L chains, this position is rather variable with >40% of sequences having a residue other than Tyr (Kabat *et al.*, 1987). Among tertiary response antibodies, two which express V_xOx1 have mutated this Tyr32 to Phe and do not belong to the group expressing highest affinity for the hapten (Berek *et al.*, 1987). Besides NQ10/12.5, other immunoglobulins or light chain dimers of known three-dimensional structure, D1.3 (Amit *et al.*, 1986), HyHEL-5 (Sheriff *et al.*, 1987), R19.9 (Lascombe *et al.*, 1989), 4-4-20 (Herron *et al.*, 1989) and Hil (Saul and Poljak, unpublished data) have a Tyr at position 32. A comparison of this region shows that in all cases V_L Tyr32 is directed outwards towards the solvent, contrary to what is seen in NQ10/12.5.

Two hot spots of expressed somatic mutation, identified by extensive sequence analysis of anti-phOx antibodies (Griffiths *et al.*, 1984), occur at V_L positions 34 and 36 (Figure 4), which are occupied by Asn and Phe residues, respectively, in NQ10/12.5. Both residues are in rather intimate contact with the hapten, as judged from the 2 $F_{obs}-F_{calc}$ map. Phe36 adopts the same conformation as that observed for Tyr36 in all other known X-ray structures. V_L residues Phe36 and Phe98, together with V_H residues Val37, Ala93 and Trp103, form the hydrophobic pocket in which the hapten is situated (Figure 3).

A difference Fourier map between the liganded and unliganded structures showed two major peaks, the largest

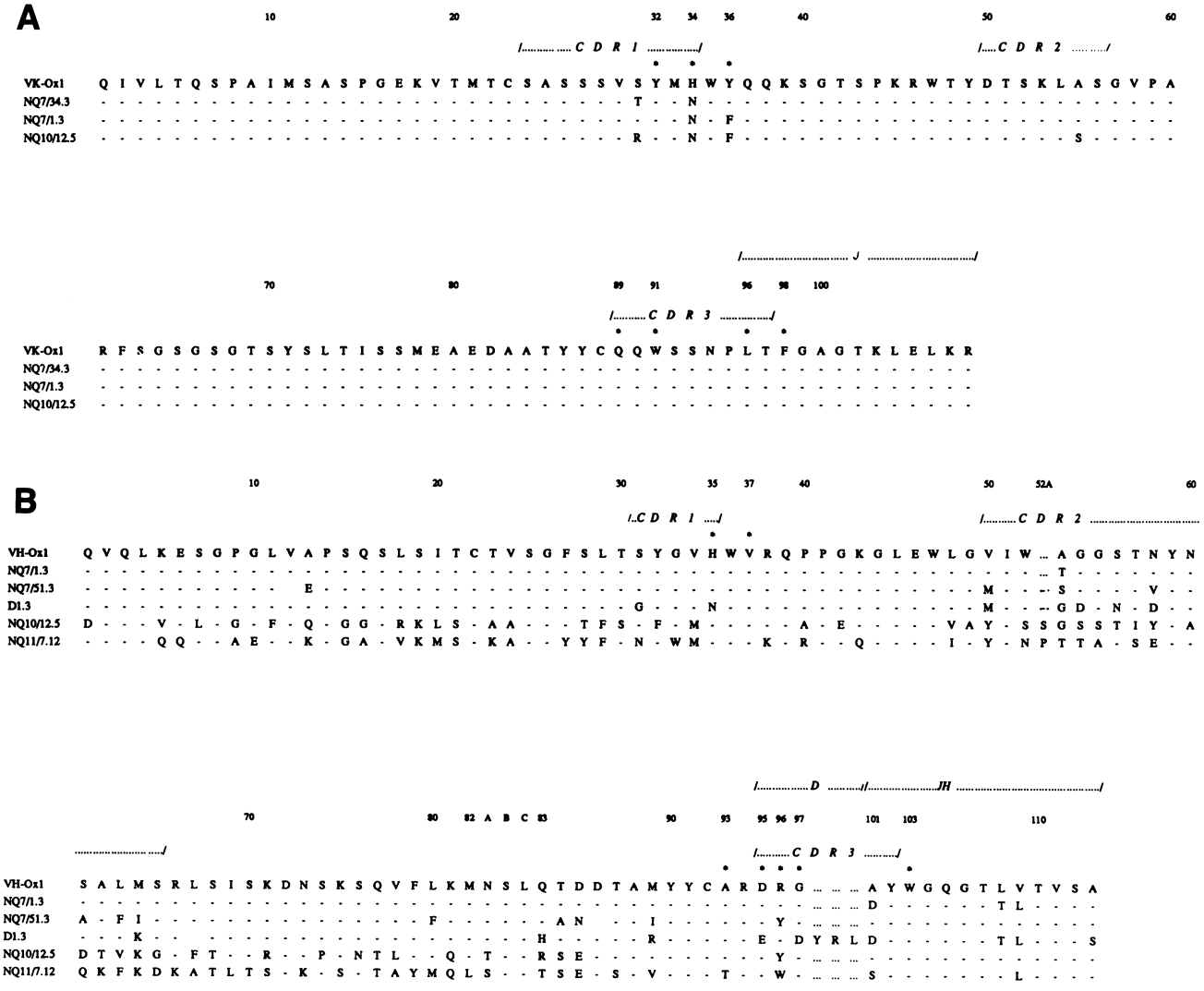

Fig. 4. Amino acid sequences of L (**A**) and H (**B**) chains V region from the anti-lysozyme antibody D1.3 (Amit *et al.*, 1986) and from selected anti-phOx hybridomas derived 14 days after primary immunization (NQ7-) and after secondary immunization (NQ10- and NQ11-). The top line corresponds in both cases to the translated sequence of the respective germline gene, V_xOx1 and V_HOx1 (Kaartinen *et al.*, 1983; Griffiths *et al.*, 1984; Berek *et al.*, 1985). Dashes in the sequences indicate identity with the top line and the positions corresponding to hapten-contacting residues in Fab NQ10/12.5 are marked with an asterisk. Gaps were introduced in V_H CDR2 and V_H CDR3 to optimize the alignment.

of which was assigned to the hapten (Figure 2). The second largest might correspond to the fixing of the side chain of V_H Tyr96 (which would in this way block the entrance to the active site), though it is not possible to exclude at this stage its assignment to the flexible aminocaproyl moiety of the hapten. Other minor peaks are seen on this map, which are consistent with small displacements of the side chains of residues V_L Tyr32, V_L Asn34 and V_H His35, all three residues are in contact with the hapten. No other detectable conformational changes in the Fab structure were observed upon hapten binding.

Among anti-phOx antibodies that use V_xOx1, there is a strong preference for the use of J_x5 probably due to an absolute requirement for a Leu residue at position 96 (Griffiths *et al.*, 1984). Other J_x segments contain Trp, Tyr or Phe rather than Leu at the homologous position. The X-ray structure of the phOx–NQ10/12.5 complex reveals that Leu96 and J region residue Phe98 (conserved in all J_x segments) interact with the hapten (Figure 3), suggesting that these restrictions conserve certain features of the amino acid sequence required to maintain the antigenic specificity of the antibody. A striking characteristic of almost all H chains associated with a V_xOx1 L chain, independently of their germline origin, is the short length of their D segment (three residues, mostly Asp-X-Gly). It seems necessary to allow a communication channel between the hapten binding pocket and the solvent region, given the unusual conformation adopted by the V_L CDR1 loop, with Tyr32 directed towards the hapten. A longer V_H CDR3 would eliminate this channel and probably partially occupy the hapten binding cavity, as suggested by superposing the structures of other known Fab fragments with a more extended D segment.

NQ10/12.5 expresses the characteristic V_xOx1 L chain of primary response antibodies, but a different H chain. Anti-lysozyme antibody D1.3, which has been refined in our laboratory at 2.5 Å resolution, uses a V_H region closely related (89% homology) to the dominant germline V_HOx1 sequence (Figure 4). A structural comparison of the two V_H subunits showed a close similarity, the rms distance of all Cα atoms (with the exception of the second and third CDRs) being 0.83 Å. The most important difference resides in the fact that D1.3 expresses a different D–J_H combination, resulting in a longer V_H CDR3. The V_H–V_L interfaces are similar, and the surface area occluded by dimerization is 1300 $Å^2$ for NQ10/12.5 and 1650 $Å^2$ for D1.3; the difference is due to the longer D segment of the latter. A direct comparison of the antigen binding sites showed that, in the D1.3 model, four residues partially fill the cavity occupied by the hapten in NQ10/12.5: V_L Arg96 (instead of Leu96 in NQ10/12.5), V_L Phe98, V_H Glu95 (instead of Asp95) and V_H Leu97, which corresponds to Gly97 in NQ10/12.5.

Discussion

The anti-phOx antibody NQ10/12.5 is composed of an L chain derived from the V_xOx1 germline gene associated with a H chain belonging to V_H group 5 (according to Dildrop, 1984), instead of the usual V_HOx1 sequence (V_H group 2). Is the mode of hapten binding by NQ10/12.5 likely to be representative of that by primary response V_xOx1/V_HOx1 antibodies, or does substitution by an apparently unrelated H chain (48% homology) necessarily impose a fundamentally altered geometry on the antibody binding site? In the absence of the crystal structure of a primary response antibody, the most straightforward alternative approach is to ask whether key phOx-contacting residues of the H and L chains of NQ10/12.5 are also found in V_xOx1/V_HOx1 antibodies.

How do the NQ10/12.5 H chain and V_HOx1 sequences compare? As shown in Figure 4, NQ10/12.5 V_H CDR1 is of the same length as in V_HOx1 antibodies, but differs from the latter at residues 32 and 34. Examination of the crystal structure of the phOx–NQ10/12.5 complex, however, reveals that neither residue is directly involved in binding the hapten. In contrast, His35, which packs against the phenyl ring of the hapten, is conserved among all available V_HOx1 sequences (Berek *et al.*, 1985). Similarly, V_H FR residues Val37 and Ala93, which play a critical role in forming the hydrophobic phOx binding pocket of NQ10/12.5, are also present in V_HOx1 antibodies. On the other hand, CDR2 of the NQ10/12.5 H chain differs extensively from the V_HOx1 sequence, in addition to being one residue longer at position 52A (Figure 4); however, as for V_L CDR2, this CDR makes no contacts with the hapten and is therefore probably not functionally 'restrained'. The D–J_H region of NQ10/12.5 is very closely related to that of V_xOx1/V_HOx1 antibodies (Figure 4). In particular three out of four antigen binding residues (Asp95, Gly97 and Trp103) are preserved, as is the overall length of CDR3/FR4. The amino acid substitution at position 96 (Arg ⇒ Tyr) also occurs among certain V_xOx1/V_HOx1 antibodies, such as NQ7/15.3 (Figure 4). The postulated similarity of hapten binding between V_HOx1/V_xOx1 antibodies and NQ10/12.5 is further supported by the similarity between the pattern of mutations in the V_xOx1 L chain of NQ10/12.5 and the mature V_xOx1/V_HOx1 antibodies. Thus, V_L of NQ10/12.5 differs from the V_LOx1 (germline) sequence at CDR1 position 34 (Asn instead of His); and 36 in FR2 (Phe instead of germline-encoded Tyr residue). Two late primary response antibodies (NQ7/1.3 and NQ7/41.3) and one secondary response antibody (NQ10/2.2.5), all of which use the V_xOx1/V_HOx1 combination, also display the same Asn and Phe substitutions at positions 34 and 36, respectively, as NQ10/12.5 (Figure 4). Indeed, those two substitutions commonly occur in secondary or tertiary response antibodies and at least one of them (Asn34) is directly responsible for a 10-fold increase in affinity for hapten (Berek and Milstein, 1987). The overall length of the CDR is maintained. As can be seen in Table II, NQ10/12.5 residues 32, 34 and 36 are strongly implicated in hapten binding (distance <4 Å). Tyr32, which appears to be hydrogen-bonded to the O atom of the oxazolone ring in the NQ10/12.5 complex structure, is highly conserved among primary and mature antibodies (Kaartinen *et al.*, 1983; Griffiths *et al.*, 1984). While NQ10/12.5 has a unique Arg at CDR1 position 31, this is probably too distant from the hapten (>6 Å) to influence binding significantly. The second CDR of the NQ10/12.5 L chain differs from the V_xOx1 sequence at position 55 (Ser instead of Ala), but this CDR makes no contacts with the hapten, although interactions with the complete hapten–carrier conjugate cannot be excluded. The CDR3 sequences of V_L NQ10/12.5 and V_xOx1, including contacting residues Gln89, Trp91 and Leu96, are completely identical. Thus, the NQ10/12.5 L chain is altogether typical of those of V_xOx1/V_HOx1 mature antibodies, even though it is associated with a MOPC21-like H chain.

The conservation of key contact residues argues that the hapten binding mode of NQ10/12.5 does not differ substantially from that of primary response V_xOx1/V_HOx1 antibodies, despite the utilization of a V_H region only 48% homologous to V_HOx1. In addition, as pointed out above, the overall three-dimensional structure of the V_H subunit of Fab NQ10/12.5 is very similar to that of Fab D1.3 (Amit *et al.*, 1986), whose V_H region shares 89% homology with V_HOx1. Thus, the NQ10/12.5 and Ox1 H chains, though encoded by apparently unrelated V_H gene segments, may in fact be functionally interchangeable due to the maintenance of important antigen binding sequences, (especially at the V_H-D-J_H junctions), as well as of general three-dimensional structure.

The V_xOx1 gene family contains between 20 and 50 related members, of which 14 have been fully sequenced (Even *et al.*, 1985). Are there obvious structural features which could explain the selective use of the V_xOx1 gene, as opposed to other members of the family? Although they are all very similar, they differ from V_xOx1 in critical contact residues. Leaving aside the two pseudogenes which contain chain termination signals in the coding region, there are only five which contain Trp91. All others contain Arg or Asn at that position. Of those five, three contain two extra residues in the essential CDR1 loop between positions 30 and 31. The two others are H13 and H6. The first, although extremely similar to V_xOx1, contains a Tyr instead of His at position 34, which may be too bulky for the tight binding pocket. As for H6, it differs in other important residues, particularly four in CDR3 (residues 92, 93, 94 and 95). Residue 94 is Pro instead of Asn, and residue 95 is Leu instead of Pro. These changes are likely to affect the orientation of the two essential residues in the Jκ5 segment, namely Leu96 and Phe98.

How heterogeneous is the anti-phOx response in terms of the number of different ways this hapten can be 'seen' by the immune system? While a definitive answer requires knowing the three-dimensional structure of at least one representative of each V_x/V_H combination, a minimum estimate can be obtained by examining the available sequence data in the context of the NQ10/12.5 model. Berek *et al.* (1985) observed that all anti-phOx antibodies that use V_xOx1 have the characteristic Asp-X-Gly D segment (see above), even though the V_H regions may originate from a variety of germ line genes. These include not only group 2 (e.g. V_HOx1) and group 5 (V_HMOPC21-like, e.g. NQ10/12.5) but also group 1 (V_HJ558-like) gene families (Berek and Milstein, 1987). Moreover, the overall length of CDR3/FR4 is constant, despite the utilization of J_H2, J_H3 and J_H4 gene segments. Other putative hapten binding residues are generally conserved as well; for example, the J558-like V_H region of NQ11/7.12 (Figure 4) has the same residues at positions 35 (His), 37 (Val) and 93 (Ala) as NQ10/12.5 and V_HOx1 (Berek *et al.*, 1985). On the other hand, many secondary response antibodies express a V_x-45.1-encoded L chain usually associated with a H chain belonging to the TEPC15 V_H gene family (V_H group 7; Dildrop, 1984). These antibodies have closely related $D-J_H$ region sequences which are systematically longer than those of V_HOx1 antibodies by four to five residues. In addition, V_L CDR1 of V_x-45.1 antibodies is six residues longer than that of V_xOx1 antibodies and there are sequence differences at most positions in contact with hapten in the NQ10/12.5 model (e.g. His35 ⇒ Glu, Gln89 ⇒ Phe). Such marked differences suggest an alternate scheme of hapten contacts in the case of V_x-45.1 anti-phOx antibodies and this conclusion is supported by the kinetics of hapten binding (Foote and Milstein, in preparation). Between these two extremes of V region usage are a minority of antibodies with V_x-ars or V_xOx related L chains associated with V_HOx1 or other H chains; at least some of these probably recognize phOx in ways other than V_xOx1 or V_x-45.1 antibodies. Thus, the BALB/c immune system appears able to 'see' this hapten by probably not very many more than two principal mechanisms.

A similar conclusion in another model system was reached by Padlan *et al.* (1985), who examined the amino acid sequences of a number of murine anti-phosphorylcholine antibodies (Gearhart *et al.*, 1981) in the light of the known three-dimensional structure of the complex between one of these (McPC603) and phosphorylcholine (Segal *et al.*, 1974). While all H chains are derived from the same V_H germline gene, the L chain sequences belong to three different classes. Nevertheless, the key V_L CDR3 contact residues, Tyr94 and Leu96, are invariant so that the overall architecture of the binding cavity probably remains unaltered.

Sequence analysis of a large number of V_xOx1 L chains revealed a very tight clustering of amino acid changes in and around CDR1 as the result of somatic point mutation of the corresponding germline gene (Griffiths *et al.*, 1984; Even *et al.*, 1985). Most affected are positions 34 (His ⇒ Asn, Gln) and 36 (Tyr ⇒ Phe, His), with scattered mutations also occurring at positions 29, 31 and 33. Are these simply mutational hot spots, or do they affect residues important in antigen binding? The L chain of NQ10/12.5 displays amino acid changes at positions 31 (Ser ⇒ Arg), 34 (His ⇒ Asn), 36 (Tyr ⇒ Phe) and 55 (Ala ⇒ Ser). As shown in Table II, residues 34 and 36 make direct contacts with the hapten and so changes here might be expected to alter binding affinity. Indeed, mutations at positions 34 and 36 are highly correlated to each other (i.e. they occur together in 20 of 24 sequences analysed, Berek and Milstein, 1987) and their appearance during the late primary response correlates well with increased antibody affinity (Griffiths *et al.*, 1984). In particular, the single substitution of V_x 34 (His ⇒ Asn) or (His ⇒ Gln) results in a 10- or 5-fold increase, respectively, in affinity. The substitution V_x 36 (Tyr ⇒ Phe) is not so obviously correlated with an affinity increase (Berek and Milstein, 1987). The strong bias towards mutation of hapten-contacting versus non-contacting residues observed among anti-phOx L chains supports the idea that antigen favours the selective expansion of clones producing antibodies bearing mutations which impart higher affinity. It should be noted, however, that only a minority of L chain contacting residues (two out of seven) are affected by somatic mutation. In some cases at least, this probably reflects structural constraints on these residues (e.g. Leu96; see above) with regard to hapten binding so that almost any substitution results in decreased affinity. It is also possible that the background mutation rate at these particular positions is simply not high enough to produce variants at a frequency which can be detected under the experimental conditions employed (sample size, immunization schedule, etc.)

A strikingly different pattern of somatic mutation is observed among V_HOx1 H chains. In this case, *none* of the observed amino acid changes unambiguously attributable to

somatic mutation involve putative hapten-contacting residues identified by extrapolation from the NQ10/12.5 model; instead, the distribution of somatic mutations is rather scattered. Marked sequence variability does, however, occur at position 101 at the $D-J_H$ junction and in the middle residue of the D segment (position 96; Figure 4). While variability at position 101 may be 'tolerated' because an antigen-contacting residue is not involved, Tyr96 interacts with the hapten in the phOx−NQ10/12.5 model (Table II). In one other case (Asp95 ⇒ Glu in NQ7/24.6) a hapten binding residue is also observed to vary. Variability at these three positions, however, is unlikely to be due to somatic mutations, but rather to the use of several different D and J_H segments and/or to imprecise joining at the V_HD and $D-J_H$ junctions (Griffiths *et al.*, 1984). Thus, in contrast to the situation for V_xOx1 L chains, sequence variability among V_HOx1 H chains clearly arising from somatic point mutation does not appear functionally significant, but rather simply to reflect the background mutation rate, although some indirect influence on antibody affinity cannot be excluded on the basis of the available binding data. Possible reasons for the absence of somatic mutation at antigen-interacting positions are the same as discussed above for V_xOx1 L chains.

Our results demonstrate the importance of three-dimensional structure determination of hapten−antibody complexes in the identification of functionally significant somatic mutations during affinity maturation of an immune response. While our conclusions concerning somatic mutation, the maintenance of junctional sequences, and the interchangeability of certain H (or L) chains are likely to be applicable to antibodies of other specificities, the requirement for detailed X-ray crystallographic analysis remains.

Materials and methods

Crystallization of native and liganded Fab NQ10/12.5

The Fab fragment of the anti-2-phenyloxazolone antibody NQ10/12.5 (IgG1, κ) was prepared and crystallized as previously described (Mariuzza *et al.*, 1985). Crystals are monoclinic, space group C2, with cell dimensions a = 119.1 Å, b = 78.7 Å, c = 85.9 Å, β = 137.7°.

Attempts to diffuse the hapten, 2-phenyl-5-oxazolone aminocaproate (phOx), into crystals of the unliganded Fab were unsuccessful, probably due to intermolecular contacts in the crystal that block entrance to the antigen binding site. Instead, isomorphous crystals of the Fab−phOx complex were grown by co-crystallization under the same conditions as for the native Fab.

Data collection and structure solution of unliganded FabNQ10/12.5

Intensity data were collected from six crystals on a Hilger-Watts four-circle diffractometer for 20 to 2.8 Å resolution (R_{merge} = 0.08−3.5 Å). The final data set contained 12947 independent reflections, ~95% of all possible ones, though only ~40% of the reflections between 3.0 and 2.8 Å had intensities greater than two standard deviations.

One heavy atom derivative, prepared by soaking a native crystal in 0.5 mM uranyl acetate, was collected at 6 Å resolution. The electron density map calculated with SIRAS (single isomorphous replacement with anomalous scattering) phases, figure-of-merit 0.89, clearly showed the variable and constant structural domains of the Fab molecule. An initial model was thus positioned in the unit cell by visual inspection of the map on an Evans and Sutherland PS330 graphic display system using the program FRODO (Jones, 1978). This starting model was built from the published structures of Fab J539 (Suh *et al.*, 1986) for the variable region and of Hy-HEL-5 (Sheriff *et al.*, 1987) for the constant region. The first and last three residues of each chain were deleted from this model, as were the central residues of the third CDRs of the H and L chains, i.e. a total of 21 residues. In addition, 41 side chains corresponding to amino acids in the first and second CDRs and to all other positions at which there were sequence differences were eliminated by converting their residue type to Gly or Ala.

Refinement of the unliganded Fab model

The correct assignment of the different subunits to the electron density was decided on the basis of a low resolution (20−8 Å) constrained−restrained refinement of the four different possibilities using the program CORELS (Sussman *et al.*, 1977). The constant and variable regions were treated as separate rigid units during this refinement.

The resulting model was subsequently refined by using CORELS, first considering the Fab molecule as four independent rigid bodies (V_L, C_L, V_H and C_H1 subunits) using 12−6 Å data, and then treating the whole molecule as 32 separate rigid bodies (corresponding approximately to the individual β-strands) to 4 Å resolution. This last step resulted in a significant improvement in the R value, from 38.0% to 33.9%, and was essential for the success of the high resolution refinement. Indeed, this type of concerted movement of large fractions of the molecule is difficult to obtain from a restrained refinement using individual atomic parameters as variables.

This model was then refined with 3.5 Å data using the program TNT (Tronrud *et al.*, 1987). During the course of the refinement, higher resolution data was gradually incorporated to the limit of 2.8 Å, and missing residues were introduced at those positions for which there was some indication on the $2F_o-F_c$ and ΔF maps. After alternative cycles of least-squares refinement and manual model building using the program FRODO, the model refined to an R-value of 0.18 for data between 7 and 2.8 Å resolution, without including solvent molecules.

The result of the final refinement cycle is summarized in Table I; coordinates and other relevant data will be deposited in the Brookhaven Protein Data Bank (Bernstein *et al.*, 1977). The last four C-terminal residues of the light chain (positions 211−214, numbering according to Kabat *et al.* (1987) are not seen in the electron density maps and are thus excluded from the present model. The last residue of the C_H1 subunit included in the model is Ile223 and three more residues (Glu42, Val63 and Tyr96, all in the V_H subunit), have no side chain atoms other than Cβ.

Structure determination of the phOx−FabNQ10/12.5 complex

A preliminary data set of the phOx−FabNQ10/12.5 complex in the 7.5−3.0 Å range was collected from two crystals on a Rigaku AFC5R four-circle diffractometer (R_{merge} = 0.09, 9524 independent reflections). The two-ring moiety of the hapten was modelled on the basis of a difference Fourier map calculated with phases from the unliganded Fab model. A refinement round of simulated annealing using program X-PLOR (Brunger *et al.*, 1987) was performed on this model; the R-factor is 0.19 for data between 7 and 3 Å resolution and the rms deviations from ideal values in bond lengths and angles are 0.02 Å and 4.1° respectively. No manual intervention nor further refinement was attempted at this stage on the model of the Fab−phOx complex.

Acknowledgements

We gratefully acknowledge Drs Claudia Berek and Pedro Saludjian for helpful insights and discussions and Françoise Gauthier for expert secretarial assistance.

References

Alzari,P.M., Lascombe,M.-B. and Poljak,R.J. (1988) *Annu. Rev. Immunol.*, **6**, 555−580.

Amit,A.G., Mariuzza,R.A., Phillips,S.E.V. and Poljak,R.J. (1986) *Science*, **233**, 747−753.

Amzel,L.M. and Poljak,R.J. (1979) *Annu. Rev. Biochem.*, **48**, 961−997.

Berek,C. and Milstein,C. (1987) *Immunol. Rev.*, **96**, 23−41.

Berek,C., Griffiths,G.M. and Milstein,C. (1985) *Nature*, **316**, 412−418.

Berek,C., Jarvis,J.M. and Milstein,C. (1987) *Eur. J. Immunol.*, **17**, 1121−1129.

Bernstein,F.C., Koetzle,T.F., Williams,G.J., Meyer,E.F., Brice,M.D., Rodgers,J.R., Kennard,O., Shimanouchi,T. and Tasumi,M. (1977) *J. Mol. Biol.*, **112**, 535−542.

Bothwell,A.L.M., Paskind,M., Reth,M., Imanishi-Kari,T., Rajewsky,K. and Baltimore,D. (1981) *Cell*, **24**, 625−637.

Brünger,A.T., Kuriyan,J. and Karplus,M. (1987) *Science*, **235**, 458−460.

Chang,C.-H., Short,M.T., Westholm,F.A., Stevens,F.J., Wang,B.-C., Furey,W., Solomon,A. and Schiffer,M. (1985) *Biochemistry*, **24**, 4890−4897.

Connolly,M.L. (1983) *J. Appl. Crystallogr.*, **16**, 548−558.

Crews,S., Griffin,J., Huang,H., Calame,K. and Hood,L. (1981) *Cell*, **25**, 59–66.
Davies,D.R., Sheriff,S. and Padlan,E.A. (1988) *J. Biol. Chem.*, **263**, 10541–10544.
Dildrop,R. (1984) *Immunol. Today*, **5**, 85–86.
Ely,K.R., Firca,J.R., Williams,K.J., Abola,E.E., Fenton,J.M., Schiffer,M., Panagiotopoulos,N.C. and Edmundson,A.B. (1978) *Biochemistry*, **17**, 158–167.
Even,J., Griffiths,G.M., Berek,C. and Milstein,C. (1985) *EMBO J.*, **4**, 3439–3445.
Gearhart,P.J., Johnson,N.D., Douglas,R. and Hood,L. (1981) *Nature*, **291**, 29–34.
Griffiths,G.M., Berek,C., Kaartinen,M. and Milstein,C. (1984) *Nature*, **312**, 271–275.
Herron,J.N., He,X.-M., Mason,M.L., Voss,E.W. and Edmundson,A.B. (1989) *Proteins*, **5**, 271–280.
Jack,R.S., Imanishi-Kari,T. and Rajewsky,K. (1977) *Eur. J. Immunol.*, **7**, 559–565.
Jones,T.A. (1978) *J. Appl. Crystallogr.*, **11**, 268–272.
Kaartinen,M., Griffiths,G.M., Markham,A.F. and Milstein,C. (1983) *Nature*, **304**, 320–324.
Kabat,E.A., Wu,T.T., Reid-Miller,M., Perry,H.M. and Gottesman,K.S. (eds) (1987) *Sequences of Proteins of Immunological Interest*. 4th edn. NIH, Bethesda, MD.
Lascombe,M.-B., Alzari,P.M., Boulot,G., Saludjian,P., Tougard,P., Berek,C., Haba,S., Rosen,E.M., Nisonoff,A. and Poljak,R.J. (1989) *Proc. Natl. Acad. Sci. USA*, **86**, 607–611.
Manser,T., Wysocki,L.J., Margolies,M.N. and Gefter,M.L. (1987) *Immunol. Rev.*, **96**, 141–162.
Mariuzza,R.A., Boulot,G., Guillon,V., Poljak,R.J., Berek,C., Jarvis,J.M. and Milstein,C. (1985) *J. Biol. Chem.*, **260**, 10268–10270.
Padlan,E.A., Cohen,G.H. and Davies,D.R. (1985) *Ann. Inst. Pasteur/Immunol.*, **136C**, 271–276.
Padlan,E.A., Silverton,E.W., Sheriff,S., Cohen,G.H., Smith-Gill,S.J. and Davies,D.R. (1989) *Proc. Natl. Acad. Sci. USA*, **86**, 5938–5942.
Reth,M., Hammerling,G.J. and Rajewsky,K. (1978) *Eur. J. Immunol.*, **8**, 393–400.
Sablitzky,F., Wildner,G. and Rajewsky,K. (1985) *EMBO J.*, **4**, 345–350.
Segal,D.M., Padlan,E.A., Cohen,G.H., Rudikoff,S., Potter,M. and Davies,D.R. (1974) *Proc. Natl. Acad. Sci. USA*, **71**, 4298–4302.
Sheriff,S., Silverton,E.W., Padlan,E.A., Cohen,G.H., Smith-Gill,S.J., Finzel,B.C. and Davies,D.R. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 8075–8079.
Suh,S.W., Bhat,T.N., Navia,M.A., Cohen,G.H., Rao,D.N., Rudikoff,S. and Davies,D.R. (1986) *Proteins*, **1**, 74–80.
Sussman,J.L., Holbrook,S.R., Church,G.M. and Kim,S.-H. (1977) *Acta Crystallogr.*, **A33**, 800–804.
Tonegawa,S. (1983) *Nature*, **302**, 575–581.
Tronrud,D.E., Ten Eyck,L.F. and Matthews,B.W. (1987) *Acta Crystallogr.*, **A43**, 489–501.
Wysocki,L., Manser,T. and Gefter,M.L. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 1847–1851.

Received on August 29, 1990